Advances in Methodology and Statistics, 2025, *22*(1), 23–30. https://doi.org/10.51936/opgo1829

A note on Korn and Simon's measure of explained variation for survival models

Janez Stare*, Ana Perišić

[a]*University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia*
[b]*University of Split, Faculty of Science, Split, Croatia*

Abstract

Among many proposals to estimate proportion of explained variation for survival data, Korn and Simon's measure is the only one actually looking at the variation of time. While this should be an advantage, its usage, as measured by citations, suggests otherwise. We look at some possible reasons for this. We find one, serious, but easily corrected, deficiency of the measure, and another property, which is natural, but not very practical when there are censored data.

Keywords: survival analysis, explained variation, Korn and Simon's measure

1. Introduction

While there were quite some papers proposing measures of explained variation for survival data (Graf et al., 1999; Harrell et al., 1982; Kent & O'Quigley, 1988; Nagelkerke, 1991; Schemper & Henderson, 2000), only two are often used in practical applications (Harrell et al., 1982; Nagelkerke, 1991). These would be fine, if those two didn't come with drawbacks that should actually make them much less appealing (Henderson et al., 2024).

More importantly, except for the proposal by Korn and Simon (1990), all others are explaining variation of something else than time (e.g., distances between survival curves, variations of ranks, or covariates). However we want to look at survival models, time to event is the outcome variable and it is always the variation of the outcome variable that one wants to understand. This fact alone should make the Korn and Simon's measure a favourite over others. But its usage says otherwise.

For example, of the four comparisons of measures of explained variation in survival analysis (Austin et al., 2015; Hielscher et al., 2010; Rahman et al., 2017; Schemper & Stare, 1996) only the first one, by far the oldest, included the Korn and Simon measure. Another illustration is the number of citations. Table 1 gives these numbers for five measures.

Before suggesting reasons for this, we first remind the reader what the Korn and Simon's measure is.

*Corresponding author
Email addresses: janez.stare@mf.uni-lj.si (JS), ana.perisic@pmfst.hr (AP)
ORCID iDs: 0000-0002-2564-8781 (JS), 0000-0001-9180-0270 (AP)

Table 1. Number of citations of five measures in the last 10 years.

Measure	No. of citations
Nagelkerke (1991)	2473
Harrell et al. (1982)	1786
Graf et al. (1999)	487
Schemper and Henderson (2000)	73
Korn and Simon (1990)	37

2. The Korn and Simon's measure

The paper from 1990 discusses two classes of measures. One involves rank correlations between observed and predicted survival times and they are not the subject of this paper. The other involves loss functions $L(y, p)$ that measure loss from predicting p and observing y. They discuss several loss functions, here we are only looking at the squared error loss, so that $L(y, p) = (y - p)^2$. But what we say about this loss function, is essentially the same for any other.

For the squared error loss the optimal predictor is the mean $E(T)$ and the expected loss (or risk) for a survival curve S, under the assumption of the correctly specified model, is

$$R\left[S\left(\cdot\right)\right] = \int \left(t - E\left(T\right)\right)^2 dF(t),$$

where $F(t) = 1 - S(t)$.

Explained variation is a proportion of the variation under the null model explained by the model. Using the above notation Korn and Simon define the explained variation as

$$EV_{KS} = \frac{R[S_0(\cdot)] - \frac{1}{n}\sum_{i=1}^{n} R[S(\cdot|x_i)]}{R[S_0(\cdot)]},$$

where $S_0(\cdot)$ denotes the null model. They define the null model to be

$$S_0(t) = \frac{1}{n}\sum_{i=1}^{n} S(t|x_i),$$

where x_i are fixed covariates. In other words, their null model is the average of the predicted survival curves.

Three distinguishing features of their approach to calculating explained variation are:

1. They assume the model to be correct and calculate the null model and the residual variances under this assumption.
2. The null model is defined as the average of the predicted survival curves.
3. Their measure is sensitive to monotonic transformations of time. As natural as this may be, this property distinguishes their measure from others.

We discuss the above in the next section.

3. Possible drawbacks of the Korn and Simon's measure

3.1. The model is assumed to be correct

To estimate the Korn and Simon measure we need the overall variance under the null model and the residual variance under the fitted model. We discuss the null model in the next subsection, here we look at the residual variance.

To estimate the residual variance, we need to look at the differences between the observed and predicted values for each subject. Korn and Simon don't do that, they estimate the predicted survival curves and use those to estimate the residual variances. This means that the actual values of times are not used after fitting the model and that censoring then represents no problem.

It is probably obvious that this is ok, but we nevertheless explain here the general idea in a few lines.

When we fit a certain model, we get estimates of the means and the residual variances for given values of covariates. Denote the estimated variance by $\hat{\sigma}^2$. Since $\hat{\sigma}^2$ is a consistent estimator of σ^2, we have

$$E(\hat{\sigma}^2) = \sigma^2.$$

Then we generate data with the fitted model and this variance. This gives us the estimate of the variance, say

$$\hat{\hat{\sigma}}^2.$$

And of course, under the assumption of consistency, we have

$$E(\hat{\hat{\sigma}}^2) = \hat{\sigma}^2$$

and

$$E\left(E\left(\hat{\hat{\sigma}}^2\right)\right) = E\left(\hat{\sigma}^2\right) = \sigma^2.$$

We illustrate the above using linear regression. We hope the graphs are self-explanatory, here are short descriptions. Figure 1 presents a scatter plot in which the points clearly do not follow a straight-line relationship. Figure 2 shows the same data with an appropriately fitted. In contrast, Figure 3 displays the same points with an incorrect linear fit. Figure 4 plots this linear fit together with points scattered around it using the variance estimated from Figure 3, demonstrating that the residual variance remains essentially the same (up to sampling error) as in Figure 3. Finally, Figure 5 explicitly illustrates this result through a simulation based on 10,000 repetitions with a sample size of 200.

So there is essentially no difference and one can safely use the fitted model to estimate the residual variance. This avoids any problems with the censored data.

3.2. The null model

To illustrate that their proposal for the null model is not a good choice, we generated data with one binary covariate and times with the Weibull distribution

$$h(t) = \gamma\lambda t^{\gamma-1} \quad \text{and} \quad S(t) = e^{-\gamma t^{\gamma}},$$

where we chose $\lambda_1 = 1$ and $\lambda_2 = 0.5$ for the two groups and $\gamma = 10$. Sample size was 1000.

Then we fitted an exponential model and calculated their measure according to their proposal, so with the null model being the average of the two predicted survival curves. We obtain $EV_{KS} = 0.118$.

The question is if taking the average of predicted survival functions as the null model makes sense. A different model will give us a different null model. Most statisticians would actually calculate the Kaplan–Meier estimator of the survival curve as the best one can do without considering covariates. And comparing the exponential fit to the KM curve will give us $EV_{KS} = -8.231$! As ridiculous as this sounds, it simply tells us that the variance of

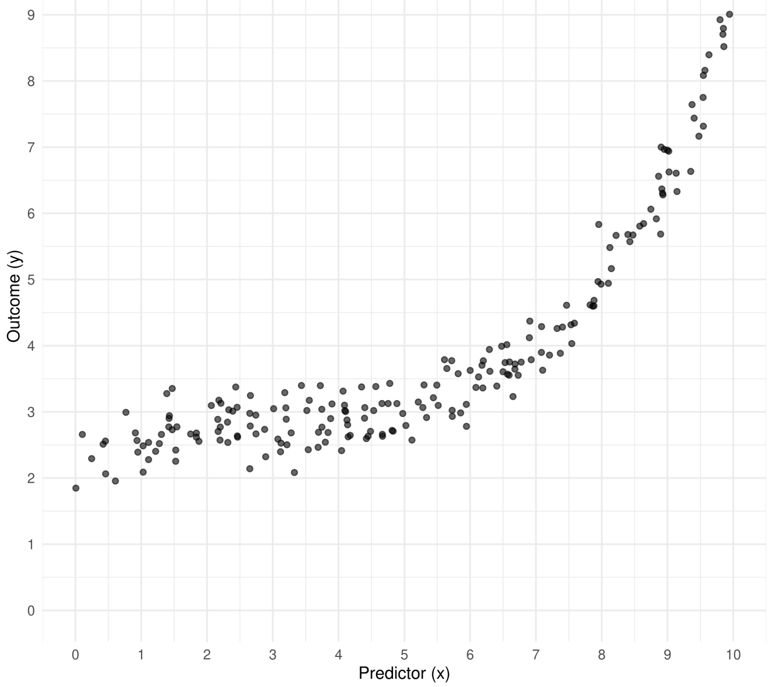

Figure 1. Cubic relationship between predictor and outcome.

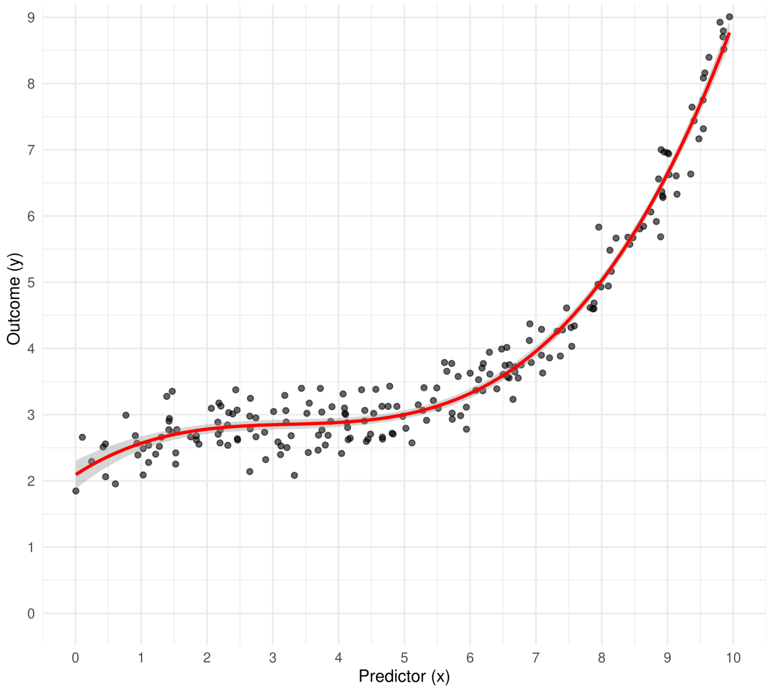

Figure 2. Best fit using a cubic polynomial.

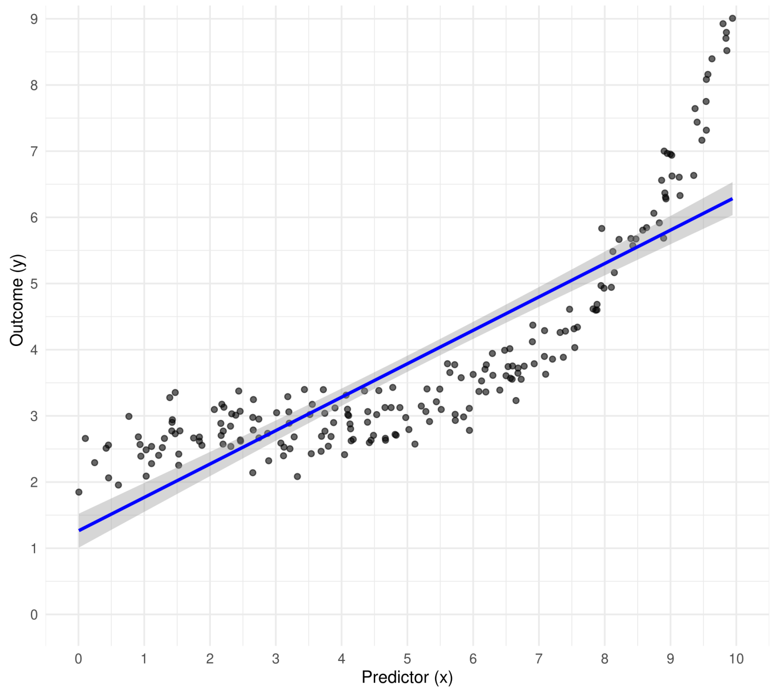

Figure 3. Wrong fit using a linear model.

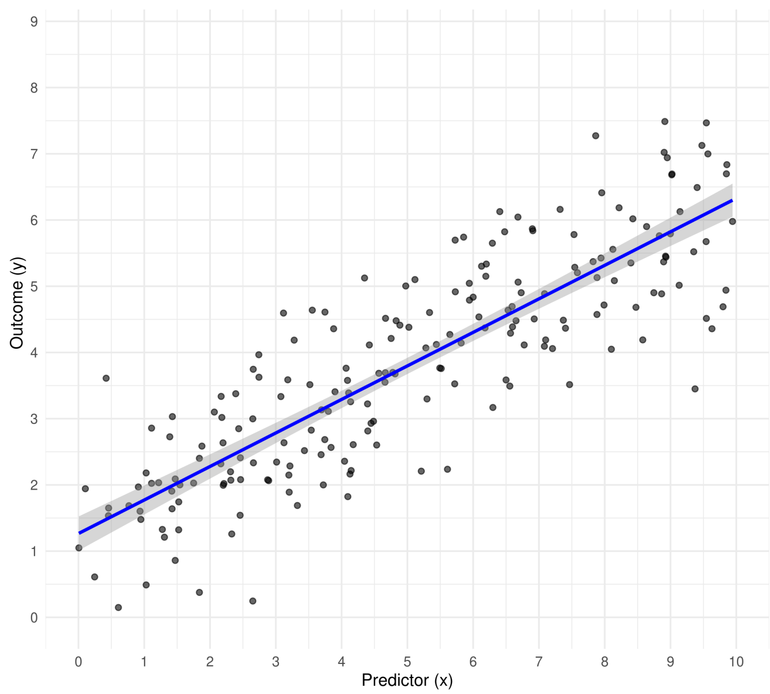

Figure 4. Data generated under the linear model with estimated variance under the wrong model.

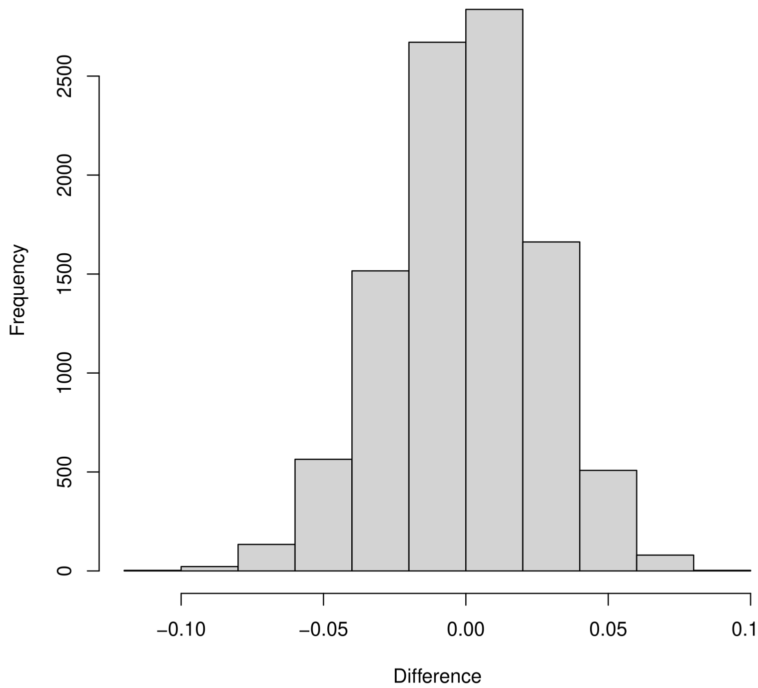

Figure 5. Distribution of the differences between R^2s calculated on original and generated data.

predictions under the exponential model is around nine times larger that the variance of predictions under the KM model.

If we fit the correct model to fit the data, so the Weibull model, we obtain $EV_{KS} = 0.880$ with their null model and $EV_{KS} = 0.870$ with KM as the null model.

The above example illustrates two situations: the first, rather extreme, where the fitted model is completely wrong, and the second, where the model is correct. Most of the time the models used in applications will be somewhere in between. But the point remains, two different fitted models will give two different null models, which doesn't make much sense (Figure 6).

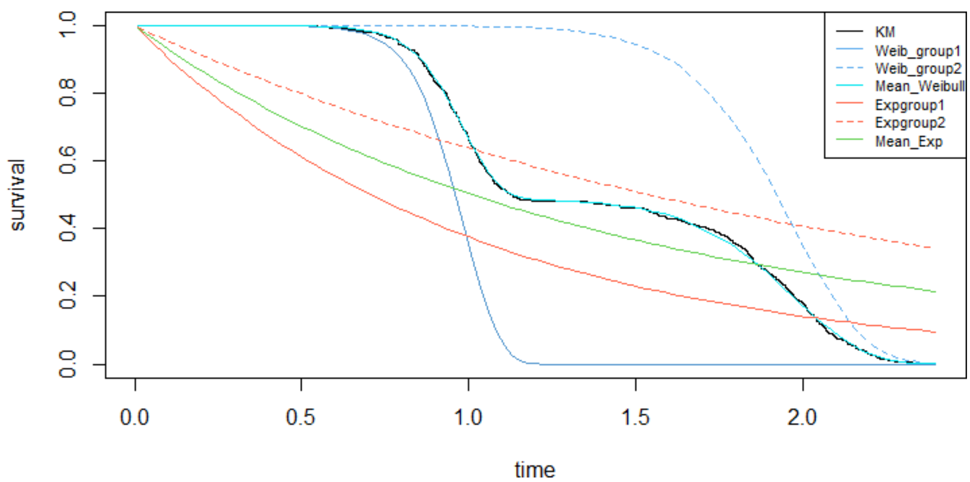

Figure 6. Illustration of the effect of using Korn and Simon's null model.

3.3. Sensitivity to monotonic transformations of time

Say clinical trials are run in Ljubljana and in New York with time to an event being a variable of interest. And say that both trials give the same result, so the same regression coefficient for the variable denoting the trial group. This is all that medical science needs to know, so one would expect that statistics are the same in both centers. But, it may well be that patients in New York survive better overall because of other aspects of treatment. This means that the baseline hazard is different between the two centers, and then the Korn and Simon measure will be different for the two models. Other measures, i.e., c-index, Schemper–Henderson, or O'Quigley & Flandre, will be the same, since they are not sensitive to monotone transformations of time.

For example, a data set with exponentially distributed times in two groups gives us $EV_{KS} = 0.30$, but after we transform times with $t_1 = 3\sqrt{t}$ we get the same results for the Cox model (apart for the baseline hazard), but the Korn and Simon measure is now 0.44.

A statistician can hardly see this property as bad, but with the Korn and Simon measure, this has an important consequence.

If data are censored after a certain T_0 then any measure of explained variation will give a different value than for non-censored data. Exceptions are those measures which inherently assume that the model is valid everywhere. Such measures are called independent of censoring, but one can make other measures also independent of censoring by extending the data under the fitted model (Kejžar et al., 2016). Unfortunately, Korn and Simon measure is an exception, unless the model is fully parametric. We cannot, for example, extend the data under the Cox model without assuming some baseline hazard. And the measure will be different for different baselines.

This is an effect of censoring that cannot be avoided with the Korn and Simon measure.

4. Conclusion

We strongly believe that the null model must not depend on the model fitted and should be the best model we can use when we are not using any covariates. The Kaplan–Meier function is an obvious choice. This is easily incorporated into the Korn and Simon measure and the measure should only be used with such a null model.

On the question of sensitivity to monotone transformations of time we are reluctant to give a strong opinion, but one has to accept that censoring beyond certain T_0 is common in survival analysis.

Funding

The authors (JS) acknowledge the financial support from the Slovenian Research and Innovation Agency under the core research programme Methodology for data analysisi in medical sciences (P3-0154).

References

Austin, P. C., Pencinca, M. J., & Steyerberg, E. W. (2015). Predictive accuracy of novel risk factors and markers: A simulation study of the sensitivity of different performance measures for the Cox proportional hazards regression model. *Statistical Methods in Medical Research*, *26*(3), 1053–1077. https://doi.org/10.1177/0962280214567141

Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assesment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, *29*(17–18), 2529–2545. https://doi.org/bg5khm

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *JAMA*, *247*(18), 2543–2546. https://doi.org/10.1001/jama.1982.03320430047030

Henderson, R., Manevski, D., Košuta, T., & Stare, J. (2024). A cautionary note on two measures of explained variation in survival analysis. *Advances in Methodology and Statistics*, *21*(1), 1–6. https://doi.org/10.51936/qxeu3315

Hielscher, T., Zucknick, M., Werft, W., & Benner, A. (2010). On the prognostic value of survival models with application to gene expression signatures. *Statistics in Medicine*, *29*(7–8), 818–829. https://doi.org/10.1002/sim.3768

Kejžar, N., Maucourt-Boulch, D., & Stare, J. (2016). On bias of measures of explained variation for survival data. *Statistics in Medicine*, *35*, 877–882. https://doi.org/10.1002/sim.6749

Kent, J. T., & O'Quigley, J. (1988). Measures of dependence for censored survival data. *Biometrika*, *75*(3), 525–534. https://doi.org/10.1093/biomet/75.3.525

Korn, E. L., & Simon, R. (1990). Measures of explained variation for survival data. *Statistics in Medicine*, *9*(5), 487–503. https://doi.org/10.1002/sim.4780090503

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691–692. https://doi.org/10.1093/biomet/78.3.691

Rahman, M. S., Ambler, G., Choodari-Oskooei, B., & Omar, R. Z. (2017). Review and evaluation of performance measures for survival prediction models in external validation settings. *Medical Research Methodology*, *17*(60). https://doi.org/10.1186/s12874-017-0336-2

Schemper, M., & Henderson, R. (2000). Predictive accuracy and explained variation in Cox regression. *Biometrics*, *56*(1), 249–255. https://doi.org/10.1111/j.0006-341X.2000.00249.x

Schemper, M., & Stare, J. (1996). Explained variation in survival analysis. *Statistics in Medicine*, *15*(19), 1999–2012. https://doi.org/dbtgzw